Mean-Field Models in neuroscience

Bruno Cessac

Nhemacs meeting
INRIA
7-05-2014

B. CESSAC

Mean-field methods in neuroscience: An introduction.

Content.

I] Introduction

Neuronal dynamics results from the complex non linear interactions between a huge number of neurons with different biophysical characteristics (neurons populations), occurring in a highly structured system, the nervous system, resulting from billions of years of evolution. Neuronal networks are the locus of a multi-scale, spatio-temporal dynamics, spanning a huge range of spatial (from molecular to brain scale) and temporal (from milliseconds to years) scales.

This is therefore a privileged field to apply a program similar to what physics has done with physical systems when deriving from the fundamental laws of physics, occurring at the microscopic level (particles), phenomenological laws characterizing the behavior of matter at higher spatio-temporal scales (mesoscopic and macroscopic). Thermodynamics, electromagnetism, fluid mechanisms, non equilibrium statistical physics are prominent examples of this, yet unachieved, program in physics.

A similar program is already ongoing in theoretical neuroscience, with some prominent achievements. The Hodgkin - Huxley equations constitute a famous example, although they do not constitute, per se, a breakthrough with the "physics program" mentioned above: Hodgkin-Huxley equations have been derived from a thorough modelling of biophysics at the molecular scale (ionic channels). In the same way, the kinetic equations modelling the synaptic transmission between neurons is fairly well described by the laws of physics.

However, when attempting to apply the same program for the description, e.g., of neuronal networks or neural masses, by a set of phenomenological equations resulting from a "suitable" averaging of neuronal dynamics ("microscopic dynamics") one is rapidly faced with prominent differences with physics. For example:

- Interactions between neurons (chemical synapses) are not symmetric;
- These interactions evolve in time depending on pre- and post-synaptic neuron activity (synaptic plasticity).
- The "graph of interactions" structure widely change depending from the nervous system part under consideration (e.g retina versus cortex).
- There is no such principle as energy or momentum conservation.
- There is no "natural" probability distribution, such as the Boltzmann - Gibbs distribution $\frac{1}{Z} e^{-E/kT}$, to characterize the statistics of events (although the concept of Gibbs distribution properly extended to handle non stationary dynamics could be useful).

Despite these differences, it is possible to adapt methods from physics to the study of neuronal systems at a mesoscopic scale. Mean-field methods constitute a prominent example of this. In this lecture we shall give an illustration of Mean-field methods applied to a famous neuronal network model, the Amari-Wilson Cowan model. After a reminder of mean-field method in statistical physics (illustrated with the Curie-Weiss model, in chapter II), we develop the analysis of Amari-Wilson-Cowan model with the so-called Dynamic Mean-Field Theory, introduced by Sompolinsky and coworkers in the field of neural networks. As a consequence, this lecture is widely inspired by Sompolinsky et al papers, although we obtain Dynamic Mean Field equations by a different approach. This is because the main goal here is not to derive these equations, but, instead, to analyze their dynamics and investigate their link with the neuronal system under study. ~~Another~~ Another goal of this course is to emphasize the differences with statistical physics models, and the model considered here provides a clean and didactical example. We also want to discuss the different "levels" of mean-field approach, from "naive" mean-field theory - which has produced several famous and efficient models, such as Jensen-Rit's model for cortical columns and epilepsy - to dynamic mean-field method which provides a more accurate of dynamics, especially fluctuations about the mean.

It is also important, as we show, to keep in mind that mean-field methods suffer several caveats. Specially, the inherent averaging procedure combined with a thermodynamic limit, leads to equations which miss many important aspects of the system they intend to describe. Additionnally, their interpretation (the interpretation of the averages) is controversial.

A last goal of this course is to propose an example, coming from neuroscience, where the "Mathemacs WP2" diagram, could be tested.

mesoscopic dynamics

$$
\begin{tikzcd}
Y \arrow[r, "\tilde{F}"] & Y \\
X \arrow[u, "\varphi_\varepsilon"] \arrow[r, "F"'] & X \arrow[u, "\varphi_\varepsilon"']
\end{tikzcd}
$$

change of scale

microscopic dynamics

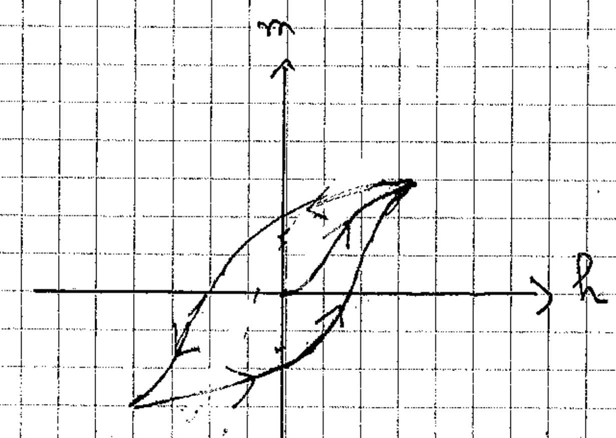
where φ_ε is an invertible mapping (homeomorphism)

II] Mean-field method in statistical physics, order parameters, phase transitions:

1) Phenomenology.

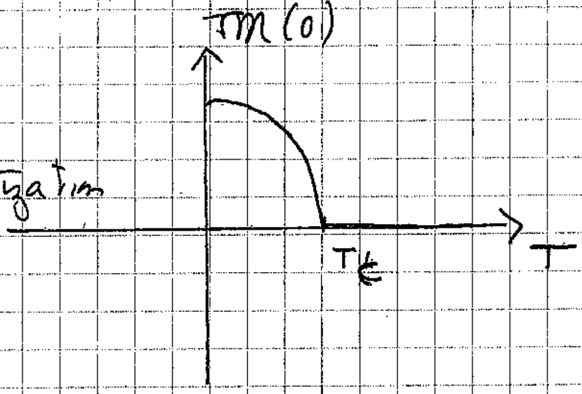
The goal here is to represent basic concepts and methods in a classical example: the Curie-Weiss model. This is a model for phase transition from paramagnetism to ferromagnetism.

Consider a piece of iron at ambient temperature. When submitted to an external magnetic field, this material acquire a spontaneous magnetization: the material remains magnetized even if the external magnetic field is removed. Moreover, magnetization exhibits hysteresis. Submitting the piece of iron to a variable external magnetic field h (constant direction but varying amplitude, with change of sign at some point) the magnetization m follows an hysteresis curve.



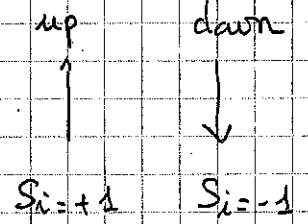
Note that magnetization has the symmetry $m(h) = -m(-h)$. Therefore, since $m(0^+) > 0$, magnetization is discontinuous at $h = 0$. The magnetization $m(0^{\pm})$ is called spontaneous magnetization.

When increasing the temperature above a critical value T_c (Curie's temperature) spontaneous magnetization disappears. The system becomes paramagnetic. When submitted to an external magnetic field the piece of iron still becomes magnetized, but this magnetization disappears when the external field is removed: $m(0) = 0$, This is an example of a (first order) phase transition.



The emergence of a mesoscopic magnetization (at the scale of the piece of iron) can be explained by phenomenological models like Ising or Curie-Weiss models, taking part of the physics but neglecting many other aspects (like dynamics).

In these models each iron atom i has a small magnetic field ("spin") S_i oriented upward or downward. The global magnetization of the piece of iron is:

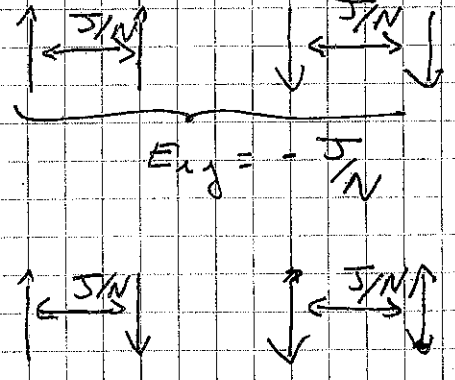


$$m = \frac{1}{N} \sum_{i=1}^{N} S_i ,$$

where N is the number of atoms.

Spins interact with each other via a ferromagnetic interaction. In the simplest examples this interaction has an intensity $J > 0$ independent of the pair of spins. The interaction energy between two spins i, j is:

$$\boxed{E_{ij} = -\frac{J}{N} S_i S_j}$$



Since $J > 0$ the energy is minimal when the two spins are aligned (up or down) and is maximal otherwise.

By extension the total energy of interaction in the system is:

Curie Weiss energy.
All spins are interacting.

$$\boxed{E = -\frac{J}{N}\sum_{1\leq i<j\leq N} S_i S_j \equiv E(\{S\})}$$

(there is no self-interaction), where $\{S\}$ denotes a spin configuration $\{S\} = (S_i)_{i=1}^{N}$.

When submitted to an external magnetic field h, each atom sees a local magnetic field h_i and the total energy of a spin configuration is:

$$\boxed{E(\{S\}) = -\frac{J}{N}\sum_{1\leq i<j\leq N} S_i S_j - \sum_{i=1}^{N} h_i S_i} \quad (1)$$

Energy likes to be minimized: the lower the energy the more stable is the corresponding configuration. This corresponds to a dynamics of energy minimization. For $h=0$ there are two global minima where all spins are aligned. In the presence of an external field (e.g. spatially homogeneous) the symmetry up/down is broken and the system chooses the configuration which minimizes E by aligning spins with the local field.

Temperature T gives however spins the freedom to violate energy minimization (spins alignment): the higher the temperature the higher the probability to violate the energy minimization rule. At $T=+\infty$ spins have up or down orientation, with probability $1/2$, independently of the configuration energy.

2. Gibbs distribution.

It is possible to construct stochastic dynamics (Markov chains) that mimics this competition between energy minimization and temperature (Glauber dynamics, Metropolis dynamics). The main constraint imposed upon these dynamics is that

Their equilibrium distribution (invariant probability of the Markov chain) is a Gibbs distribution.

$$\boxed{P_N(\{S\}) = \frac{1}{Z_N} \exp - \beta E_N(\{S\}),} \qquad (2)$$

where $\beta = \frac{1}{kT}$ is the inverse temperature (k is typically the Boltzmann constant, but here it can be set to $k=1$), Z_N is the partition function:

$$\boxed{Z_N = Z_N(\beta, h) = \sum_{\{S\}} e^{-\beta E_N(\{S\})},} \qquad (3)$$

i.e. the normalization factor ensuring that P_N is a probability distribution. In our notations we have made explicit the dependence of P and E in the system size N; we have also indicated that Z_N (via E_N) depends on temperature (β) and external field.

Note that Z_N is, additionally, the log-generating function of spins and energy cumulants.

Indeed:

$$\frac{\partial \log Z_N}{\partial h_i} = \frac{-\beta}{Z_N} \sum_{\{S\}} e^{-\beta E_N(\{S\})} \frac{\partial E_N}{\partial h_i}$$

$$= +\beta \sum_{\{S\}} P_N(\{S\})\, S_i = \beta \langle S_i \rangle$$

where $\langle S_i \rangle$ is the average value, under $P_N(\{S\})$ of the spin orientation at site i: this is the local magnetization

$m_i = \langle S_i \rangle$ (which depends on N, h, β). Therefore defining:

$$\boxed{F_N = -\frac{1}{\beta} \log Z_N} \quad (4)$$

we have:

$$\boxed{m_i = -\frac{\partial F_N}{\partial h_i}.} \quad (5)$$

Likewise:

$$\frac{\partial^2 F_N}{\partial h_i \partial h_j} = -\beta^{-1} \frac{\partial}{\partial h_j} \left[\frac{\partial}{\partial h_i} \log Z_N \right]$$

$$= \frac{\partial}{\partial h_j} \left[\frac{1}{Z_N} \sum_{\{S\}} e^{-\beta E_N(\{S\})} S_i \right]$$

$$= -\frac{1}{Z_N^2} \frac{\partial Z_N}{\partial h_j} \sum_{\{S\}} e^{-\beta E_N(\{S\})} S_i - \beta \frac{1}{Z_N} \sum_{\{S\}} e^{-\beta E_N(\{S\})} S_i \frac{\partial E_N}{\partial h_j}$$

$$= -\frac{\beta}{Z_N^2} \sum_{\{S\}} e^{-\beta E_N(\{S\})} S_j \sum_{\{S\}} e^{-\beta E_N(\{S\})} S_i + \beta \frac{1}{Z_N} \sum_{\{S\}} e^{-\beta E_N(\{S\})} S_i S_j$$

$$= \beta \left[\langle S_i S_j \rangle - \langle S_i \rangle \langle S_j \rangle \right].$$

Therefore, the pairwise spin correlation is given by:

$$(6) \quad \boxed{\langle S_i S_j \rangle - \langle S_i \rangle \langle S_j \rangle = \beta^{-1} \frac{\partial^2 F_N}{\partial h_i \partial h_j} = \beta^{-1} \frac{\partial m_i}{\partial h_j}}$$ Fluctuation response

Higher order cumulants can be generated likewise. In the same way, the average energy is given by:

$$\langle E_N \rangle = -\frac{\partial \log Z_N}{\partial \beta} \equiv \sum_{\{S\}} E_N(\{S\}) P_N(\{S\})$$

Defining the entropy

$$S[P_N] = -k \sum_{\{S\}} P_N(\{S\}) \log P_N(\{S\}),$$

we have:

$$\langle E_N \rangle - T S[P_N] = \sum_{\{S\}} \left[E_N(\{S\}) + kT \log \left(\frac{e^{-\beta E_N(\{S\})}}{Z_N} \right) \right] P_N(\{S\})$$

$$= -kT \log Z_N = F_N$$

so that F_N is the free energy.

3) Thermodynamic limit and phase transition.

This model is fine as it correctly describes the microscopic physics, but, in its actual stage, it does not explain the appearance of a spontaneous magnetization $m(0^+) > 0$. Indeed, if $h = 0$, the energy $E_N(\{S\})$ has the symmetry $E_N(\{S\}) = E_N(-\{S\})$. As a consequence the local magnetization of spin i obeys:

$$\langle S_i \rangle = \sum_{\{S\}} S_i \frac{\exp -\beta E_N(\{S\})}{Z_N} = \sum_{\{S\}} -S_i \frac{\exp -\beta E_N\{-S\}}{Z_N}$$

$$= -\langle S_i \rangle,$$

so that $\langle S_i \rangle = 0$. There is no spontaneous magnetization

On mathematical grounds, we will now show that the spontaneous magnetization is acquired when $N \to +\infty$ (thermodynamic limit). On physical grounds, this limit corresponds to assimilate the huge number of atoms (Avogadro number $\mathcal{N} = 6.02 \times 10^{23}$) with an infinite number (although huge number does not mean infinite — the discussion would require to introduce more carefully dynamics and characteristic times).

As a preliminary, but essential remark, note that "taking the thermodynamic limit" is not a straightforward affair. Indeed, it is easy to see that the energy $E_N(\{S\})$ has <u>no limit</u> for most configuration:

Take e.g. a one dimensional lattice and consider configurations such as

$$\underbrace{-1}_{1}\ \underbrace{1\,1}_{2}\ \underbrace{-1\,-1\,-1}_{3}\ \underbrace{1\,1\,1\,1}_{4}\ \dots$$

Therefore, it would make no sense to try to define objects such as " $\lim_{N \to +\infty} P_N(\{S\})$ "

There exist entire books in mathematical physics dealing with the definition of Gibbs states in the thermodynamic limit (e.g. <u>Dobrushin - Lanford - Ruelle</u> construction). In the present context it is sufficient to know that, in the Curie-Weiss model, the free energy density (free energy per spin):

$$f = \lim_{N \to +\infty} \frac{F_N}{N}$$

exists and obeys the same properties as F_N (generating function). Especially the local magnetization m_i, is still defined in the thermodynamic limit and obeys

$$m_i = -\frac{\partial f}{\partial h_i}.$$

The computation of f can be done via the computation of the finite size partition function Z_N. Let us do it. We have

$$Z_N = \sum_{\{S\}} \exp \frac{\beta J}{N} \sum_{1\leq i<j\leq N} S_i S_j \exp \beta \sum_{i=1}^{N} h_i S_i$$

For simplicity we consider here a uniform magnetic field, $h_i = h$. We set

$$\mathcal{V} = \beta J \quad ; \quad B = \beta h$$

This gives:

$$Z_N = \sum_{\{S\}} \exp \frac{\mathcal{V}}{N} \sum_{1\leq i<j\leq N} S_i S_j \exp B \sum_{i=1}^{N} S_i.$$

Note that:

$$\sum_{1\leq i<j\leq N} S_i S_j = \frac{1}{2} \sum_{i=1}^{N} \sum_{\substack{j=1\\ j\neq i}}^{N} S_i S_j = \frac{1}{2}\left(\sum_{i=1}^{N} S_i \sum_{j=1}^{N} S_j - \underbrace{\sum_{i=1}^{N} S_i^2}_{=N}\right),$$

therefore

$$Z_N = \exp -\frac{\mathcal{V}}{2} \sum_{\{S\}} \exp \frac{\mathcal{V}}{2N}\left(\sum_{i=1}^{N} S_i\right)^2 \exp B \sum_{i=1}^{N} S_i.$$

Using the identity:

$$\exp \frac{a}{2} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp -\frac{x^2}{2} \exp \sqrt{a}\, x \, dx \qquad (a \geq 0),$$

we obtain:

$$\exp \frac{\nu}{2N} \left(\sum_{i=1}^{N} S_i \right)^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp \frac{-x^2}{2} \exp \left(\sqrt{\frac{\nu}{N}} \sum_{i=1}^{N} S_i \right) x \, dx,$$

and:

$$Z_N = \frac{1}{\sqrt{2\pi}} \exp -\frac{\nu}{2} \int_{-\infty}^{+\infty} e^{-x^2/2} \sum_{\{S\}} \prod_{i=1}^{N} \exp \left(x \sqrt{\frac{\nu}{N}} + B \right) S_i \, dx.$$

We have:

$$\sum_{\{S\}} \prod_{i=1}^{N} \exp \left(x \sqrt{\frac{\nu}{N}} + B \right) S_i = \sum_{\substack{S_1 = \pm 1 \\ S_2 = \pm 1 \\ \vdots \\ S_N = \pm 1}} \prod_{i=1}^{N} \exp \left(x \sqrt{\frac{\nu}{N}} + B \right) S_i$$

$$= \prod_{i=1}^{N} \sum_{S_i = \pm 1} \exp \left(x \sqrt{\frac{\nu}{N}} + B \right) S_i = \prod_{i=1}^{N} 2 \operatorname{ch} \left(x \sqrt{\frac{\nu}{N}} + B \right) = 2^N \operatorname{ch}^N \left(x \sqrt{\frac{\nu}{N}} + B \right)$$

As a consequence, and since $\operatorname{ch}(u) > 0$:

$$Z_N = \frac{1}{\sqrt{2\pi}} \exp -\frac{\nu}{2} \int_{-\infty}^{+\infty} \exp \left[-\frac{x^2}{2} + N \log \left(2 \operatorname{ch} \left(x \sqrt{\frac{\nu}{N}} + B \right) \right) \right] dx.$$

Setting $\eta = \dfrac{x}{\sqrt{\nu N}}$ and

$$g(\eta) = \frac{\nu \eta^2}{2} - \log \left(2 \operatorname{ch}(\nu \eta + B) \right), \qquad (7)$$

we get:

$$Z_N = \sqrt{\frac{\nu N}{2\pi}} \exp -\frac{\nu}{2} \int_{-\infty}^{+\infty} \exp -N g(\eta) \, d\eta.$$

When $N \to +\infty$ this integral can be estimated using the steepest descent method:

Assume that $g(\eta)$ has a global minimum at η_0, then:

$$\lim_{N \to +\infty} \frac{1}{N} \log Z_N = g(\eta_0).$$

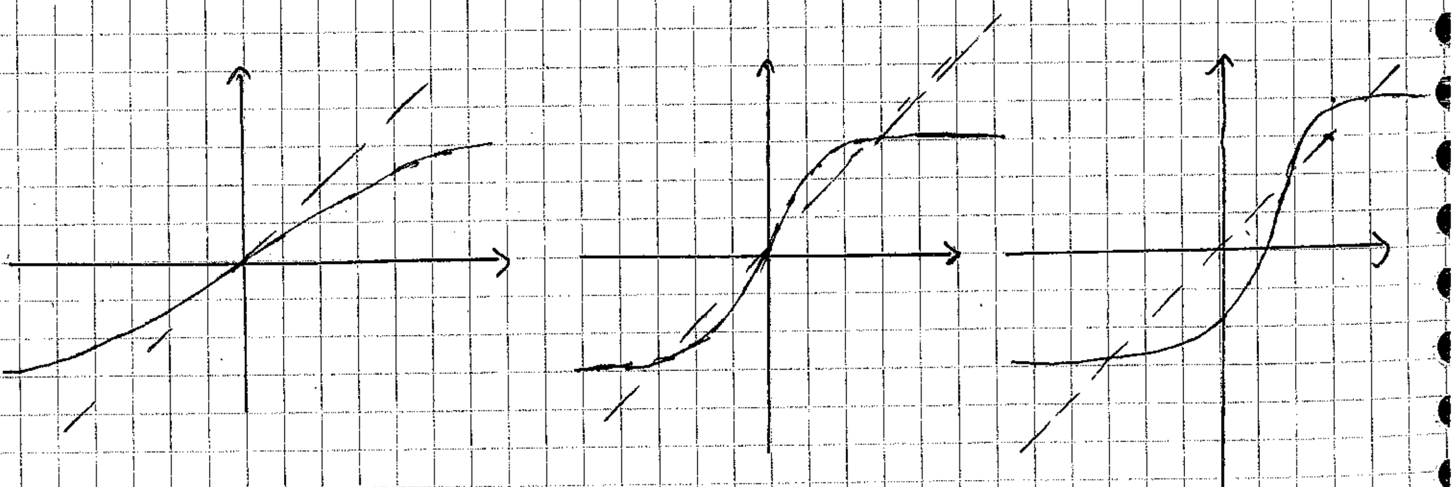
Since:

$$\frac{dg}{d\eta} = +\nu\eta - \nu \tanh(\nu\eta + B),$$

extrema are given by:

$$\boxed{\eta = \tanh(\nu\eta + B).} \qquad (8)$$

Depending on ν, B this equation can have one or several solutions. In the last case there are minima and maxima of $g(\eta)$.



$B=0, \ \nu < 1$ $\qquad$ $B=0, \ \nu > 1$ $\qquad$ $B > 0$

For $B \neq 0$ there is only one minimum. When there are several minima ($B=0$, $\nu > 1$) we have:

$$\lim_{N\to+\infty} \frac{1}{N} \log Z_N = \sum_{i=1} g(\eta_i^*) . \qquad \text{(two symmetric minima)}$$

4) <u>Interpretation - order parameter</u>

In the case of a uniform magnetic field, the global magnetization is:

$$m = \frac{1}{N} \sum_{i=1}^{N} \langle S_i \rangle = -\frac{1}{N\beta} \frac{\partial \log Z_N}{\partial h},$$

so that, in the limit $N \to +\infty$:

$$m = -\frac{1}{\beta} \frac{\partial}{\partial h} \lim_{N\to+\infty} \frac{1}{N} \log Z_N .$$

Consider first the case of a unique minimum. We have:

$$m = -\frac{1}{\beta} \frac{\partial g}{\partial h} = -\frac{1}{\beta} \frac{\partial g}{\partial B} \frac{\partial B}{\partial h} = -\frac{\partial g}{\partial B} = \tanh(\nu \eta_0 + B).$$

Therefore:

$$\boxed{m = \eta_0 = \tanh(\nu\eta_0 + B) = \tanh(\nu m + B)} \qquad (9)$$

This is the <u>equation for the magnetization</u> (first obtained by Curie, using other methods).

The spontaneous magnetization $m(0^{\pm})$ can be obtained by taking the limit $B \to 0^{\pm}$.

m characterizes the transition from paramagnetism to ferromagnetism. It is called an <u>order parameter</u>.

For $B = 0$, the transition occurs for $\nu = 1$, corresponding to:

$$\beta J = 1 = \frac{J}{k T_c} \Rightarrow \boxed{T_c = \frac{J}{k}}$$ critical temperature

5) Mean-field approximation

There is another way to obtain the mean-field equation (9) directly, by making the following approximation. Define the local field "viewed" by neuron i as:

$$\boxed{h_i^{(loc)} = \frac{J}{N} \sum_{j \neq i} S_j + h_i}, \quad (10)$$

so that the energy of spin i is $E_i = -h_i^{(loc)} S_i$, whereas the total energy is $E = +\sum_i E_i = -\sum_i h_i^{(loc)} S_i$.

Mean-field approximation consists of replacing each S_j by its mean value so that:

$$h_i^{(loc)} \sim \frac{J}{N} \sum_{j \neq i} \langle S_j \rangle + h_i = J m + h_i$$

and

$$E_i \sim -S_i (J m + h_i)$$

Then:

$$P_N(\{S\}) = \frac{1}{Z_N} \exp -\beta E_N(\{S\})$$

$$\sim \frac{1}{Z_N} \prod_{i=1}^{N} \exp \beta S_i (J m + h_i),$$

whereas:

$$Z_N = \sum_{\{S\}} \exp -\beta E_N(\{S\})$$

$$\simeq \sum_{\substack{S_1=\pm 1 \\ \vdots \\ S_N=\pm 1}} \prod_{i=1}^{N} \exp \beta S_i (Jm + h_i)$$

$$= \prod_{i=1}^{N} \sum_{S_i=\pm 1} \exp \beta S_i (Jm + h_i)$$

$$= \prod_{i=1}^{N} Z_N(i),$$

where $Z_N(i) = 2\,\mathrm{ch}(\beta(Jm+h_i))$.

Therefore:

$$P_N(\{S\}) \simeq \prod_{i=1}^{N} P_N(S_i) \quad \text{with}$$

$$P_N(S_i) = \frac{\exp \beta S_i (Jm + h_i)}{Z_N(i)}.$$

In the mean-field approximation, spins are independent.

Let us now compute the magnetization of spin i: $\langle S_i \rangle$.

$$\langle S_i \rangle = P_N(S_i = 1) - P_N(S_i = -1)$$

$$\simeq \frac{\exp \beta(Jm+h_i) - \exp -\beta(Jm+h_i)}{2\,\mathrm{ch}(\beta(Jm+h_i))}$$

$$= \tanh(\beta(Jm+h_i)).$$

In the case of an uniform magnetic field, we recover (9)

$$m = \tanh(\beta(Jm+h)).$$

The computation easily extends to the non uniform case. (Note that the exact computation of the previous section can also be extended as well to the non-uniform case).

We arrive then at the following unexpected but important conclusion.

In Curie-Weiss model, the mean-field approximation, where spins are assumed to be independent, becomes exact in the thermodynamic limit $N \to +\infty$.

This provides a straightforward way to compute the magnetization without using the technicalities of the previous section.

Unfortunately, mean-field approximation is <u>wrong</u> in models such as Ising model where connectivity is <u>local</u> (nearest neighbours) in contrast to the <u>full connectedness</u> of Curie-Weiss. especially, in Ising model, the mean-field approximation gives a wrong critical temperature.

6) <u>Extension of mean-field methods in statistical physics.</u>

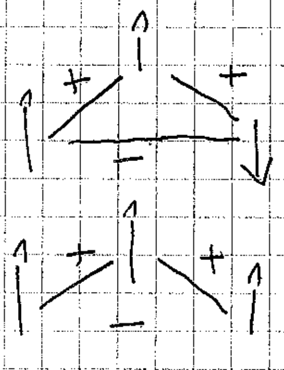
There are many domains in physics where mean-field methods give interesting results. Here we would like to briefly comment of an extension of Curie-Weiss model to the case where the spin interactions are random, symmetric and signed. This is an example of a <u>spin-glass</u> model called the Sherrington-Kirkpatrick model (78)

Thus, the energy of interaction between spins is:

$$E_N(\{S\}) = -\sum_{i<j} J_{ij} S_i S_j - \sum_{i=1}^{N} h_i S_i,$$

where $J_{ij} = J_{ji}$ and J_{ij} are i.i.d Gaussian random variables $\mathcal{N}(0, \frac{1}{N})$.

In contrast to Curie Weiss model, energy has many minima due to frustration (see figure).



Frustrated triangle.

Both configurations have a "frustrated" spin.

Moreover, the presence of disorder (random J_{ij}'s) leads to a phase diagram quite more complex than Curie-Weiss model, where magnetisation is not a sufficient order parameter. Moreover, the

Mean field theory is quite a bit harder to solve. Let us briefly figure out why. The variation of the local magnetic field h_j induces a variation of the magnetization m_j. If the variation of local field is small the variation of m_j is given by the linear response formula (6). Let us write:

$$\chi_{jj} \equiv \frac{\partial m_j}{\partial h_j} = \beta\left[\langle S_j S_j\rangle - \langle S_j\rangle\langle S_j\rangle\right].$$

In a spin glass, the definition of the local field $h_i^{(loc)}$ (eq. (9)) becomes:

$$h_i^{(loc)} = \sum_{j \neq i} J_{ij} S_j + h_i,$$

so that the field induced by spin j at site i is

$J_{ij} S_j$. Now, if one makes a slight variation of the local field $h_i^{(loc)}$, $\delta h_i^{(loc)}$, this will induce a change in the magnetisation at site j, given by $\delta m_j = \chi_{ji}\, \delta h_i^{(loc)}$. This induces a feedback effect on spin i: the variation of m_j, δm_j, induces an average variation of the local field at site i given by $J_{ij}\, \delta m_j = J_{ij}\, \chi_{ji}\, \delta h_i^{(loc)}$. Therefore, we are not free to tune the local field $h_i^{(loc)}$ as we wish. If we vary, e.g. the local external field h_i, by an amount δh_i, the total variation of the complete local field $h_i^{(loc)}$ will be

$$i \underset{J_{ji}}{\overset{J_{ij}}{\rightleftarrows}} j$$

$$\delta h_i^{(loc)} = h_i^{(loc)} + \delta h_i + J_{ij}\, \chi_{ji}\, \delta h_i .$$

Now, when we use the mean-field approximation

$$h_i^{(loc)} \sim \sum_{j \neq i} J_{ij}\, m_j + h_i ,$$

the magnetization m_j ought to contain the influence of spin i on j

$$m_j = m'_j + \chi_{jj}\, J_{ji}\, m_i ,$$

where m'_j is the magnetization in the naive mean-field approach (without feedback). In the naive mean-field approach we would have:

$$m_i = \tanh\left[\beta\left(\sum_j J_{ij}\, m'_j + h_i\right)\right]$$

$$= \tanh\left[\beta\left(\sum_j J_{ij}\, m_j + h_i - \sum_j J_{ij} J_{ji}\, \chi_{jj}\, m_i\right)\right].$$

Now $\chi_{jj} = \dfrac{\partial m_j}{\partial h_j} = \dfrac{\partial}{\partial h_j}\left[\tanh\left[\beta\left(\sum_k J_{jk}\, m'_k + h_j\right)\right]\right]$

$$= \beta\left(1 - \tanh^2\left[\beta\left(\sum_k J_{jk} m'_k + h_j\right)\right]\right)$$

$$= \beta(1 - m_j^2).$$

Therefore, finally:

$$(11) \quad \boxed{m_i = \tanh\left[\beta\left(\sum_j J_{ij} m_j + h_i - \beta \sum_j J_{ij} J_{ji} (1 - m_j^2) m_i\right)\right].}$$

This equation couple the m_j's via the feedback term.

In spin glasses J_{ij}'s are random $\mathcal{N}(0, 1/N)$ but symmetric. Therefore,

$$E\left[\sum_j J_{ij} J_{ji}\right] = \sum_j E[J_{ij}^2] = 1.$$

On the opposite, where J_{ij}'s to be independent (and asymmetric), as in the next chapter example of a neuronal network, would we have $E[\sum J_{ij} J_{ji}] = 0$, so that the feedback would play no role on average. (This has to be further justified, due to the non linearity in eq. (11); see Cessac 1995 for details).

Eq. (11) somewhat illustrates the complexity of mean-field equations. By properly taking the average over the J_{ij}'s, one ends up with the so-called TAP equations (Thouless-Anderson-Palmer). They have an infinite number of solutions ($N \to +\infty$) at low temperature.

Note that this problem does not appear in Curie-Weiss model. This is because:

$$\sum_j J_{ij}^2 = \sum_j \frac{J^2}{N^2} = \frac{J^2}{N},$$

which tends to 0 as $N \to +\infty$. There is no feedback term in Curie-Weiss model. Incidentally, this remark draws our attention on the importance of the scaling of interactions, especially their variance.

III Mean-field equations for neural networks

We would now like to extend the strategy developped in section II to the case of neural networks models. As there exist many models in the field we shall mainly focus on one example, the Amari - Wilson - Cowan (AWC) model, which has several advantages:

(i) It is well known;
(ii) It is (relatively) simple;
(iii) The mean-field approach can be developed quite far, on analytical and mathematical grounds, thanks to the seminal work of people like Amari, Sompolinsky et al, Magnot and Samuelides, Faugeras et al.
(iv) It allows one to figure out the limits of statistical physics approach and its extension; by the way it also makes explicit the caveats in mean-field approaches even when they are exact and rigorous.

1) The Amari - Wilson - Cowan model.

We will not derive these equations from biophysical models of neurons and synapses, but take them for granted.

Thus, let us consider a set of N neurons. Denote by V_i the membrane potential of neuron i. The activity of a neuron is characterized by its firing rate ν_i (probability that this neuron emits a spike between t and $t+dt$). This is non linear, sigmoidal function of neuron's membrane potential:

$$\nu_i(t) = f_g(S_i(t));$$

where typically, $f_g(x) = \dfrac{1+\tanh(gx)}{2}$ or $f_g(x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{x} e^{-u^2/2}\,du$. Although it has no direct interpretation in terms of firing rate (probability) it is also useful and illuminating to consider the case when $f_g(x) = \tanh(gx)$.

The parameter g, or 'gain', plays a crucial role as it controls the non linearity of the sigmoid (slope at the origin). It characterizes in particular the steepness of neurons's firing rate fluctuations in response to fluctuations of membrane potential near 0.

Neurons are connected by synapses. In an AWC model

The synaptic strength (weight) is characterized by a real number: J_{ij} is a synaptic weight of the connection between pre synaptic neuron (j) and post synaptic neuron (i).

It can be positive (excitation), negative (inhibition) or zero (no connection). It is non symmetric:

$$J_{ij} \neq J_{ji}$$

Neuron i can also be submitted to a local, external stimulus θ_i. We assume here that J_{ij}s do not depend on time (no plasticity)

The AWC model is characterized by the set of differential equations:

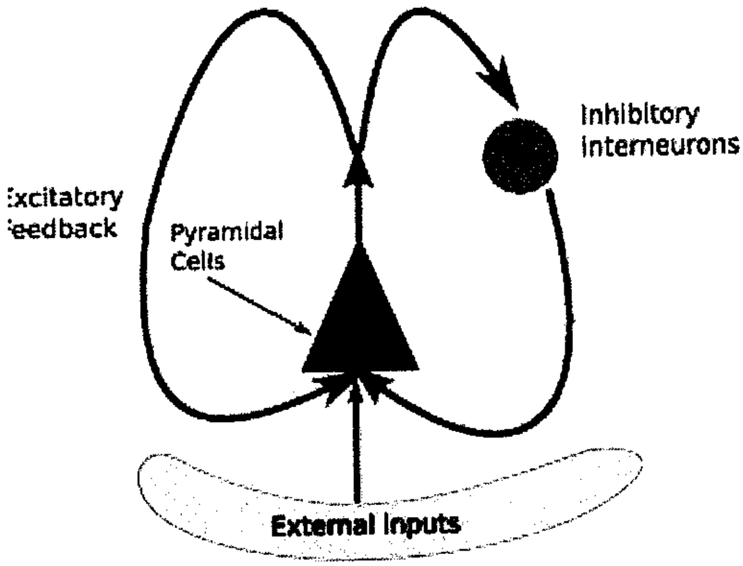
$$\frac{dV_i}{dt} = -\frac{V_i}{\tau_i} + \sum_{j=1}^{N} J_{ij} f_g(V_j(t)) + \theta_i \; ; \quad i = 1..N$$

τ_i is the characteristic leak rate of neuron i ($\tau_i = RC_i$, where C_i is the membrane capacity). For simplicity, we shall take $\tau_i = 1$. Finally, dynamics can be submitted to a white noise $\xi_i(t)$ with intensity σ. This gives, finally:

(1) $$\boxed{\frac{dV_i}{dt} = -V_i + \sum_{j=1}^{N} J_{ij} f_g(V_j(t)) + \theta_i + \sigma \xi_i(t)}$$

2) Synaptic weights

In the cortex several species of neurons exist. The connectivity between each neuron population depends on brain area and begins to be (well) known. Typically, one obtains a graph of the following type.



From a modelling point of view, one can represent this situation as follows. We have K populations of neurons indexed by a superscript α. Thus $V_i^{(\alpha)}$ is the membrane potential of i-th neuron in population α. The synaptic weight from j to i is still denoted by J_{ij}, but, to take into account variability inside connections from population β (pre-synaptic) to population α (post-synaptic), the synaptic weight from j (population β) to i (population α) is a <u>random variable</u> with mean $m_{\alpha\beta}$ and variance $\sigma_{\alpha\beta}^2$.

Since we are going to consider a mean-field approach we shall consider a fully connected network. This means that, if we call N_β the number of neurons in population β then $m_{\alpha\beta}$ is proportional to $1/N_\beta$. We set, for historical reasons (coming from spin-glasses methods applied to neuronal networks):

$$m_{\alpha\beta} = \frac{\bar{J}_{\alpha\beta}}{N_\beta}.$$

$\bar{J}_{\alpha\beta}$ can be either positive (excitatory population), negative (inhibitory population) or zero (balanced case). Likewise the variance of J_{ij}'s is:

$$\sigma^2_{\alpha\beta} = \frac{J^2_{\alpha\beta}}{N_\beta}.$$

Let us briefly comment this scaling. It ensures that the synaptic contribution to neuron i membrane potential (synaptic field) given by:

$$\boxed{G_i(t) = \sum_{j\in\beta} J_{ij}\, f(V_j(t)) + \theta_i} \qquad (2)$$

has a bounded mean and variance (with respect to J_{ij}'s distribution) as $N_\beta \to +\infty$.

Let us now add a few simplifying hypotheses.

(i) The J_{ij}'s are <u>Gaussian</u>. We shall discuss an extension of this case later on.

(ii) They are <u>independent</u>. We shall briefly discuss the <u>correlated case</u> at the end of this lecture.

(iii) As a first example we shall consider only <u>one population</u>. Extensions to <u>K populations</u> will be <u>discussed</u>.

To summarize:

J_{ij} are i.i.d. random variable, Gaussian:

$$J_{ij} \sim \mathcal{N}\left(\frac{\bar{J}}{N}, \frac{J^2}{N}\right)$$

where $\quad [\overline{J_{ij}, J_{kl}}] = \frac{J^2}{N} \delta_{ik} \delta_{kl}$.

In the simplest example we shall finally consider $\bar{J} = 0$ (<u>balanced case</u>). Then, we may write synaptic weights in the form:

$$J_{ij} = \frac{W_{ij}}{\sqrt{N}} \quad \text{with } W_{ij} \sim \mathcal{N}(0,1).$$

3) <u>Dynamic Mean-Field Theory</u> (DMFT)

We are therefore considering a dynamical system:

$$\boxed{\frac{dV_i}{dt} = -V_i + \frac{1}{\sqrt{N}} \sum_{j=1}^{N} W_{ij} f_g(V_j(t)) + \theta_i + \sigma \xi_i(t).} \quad (3)$$

This is a <u>stochastic</u>, non <u>linear</u> dynamical system, depending on <u>$N^2 + N + 1$ parameters</u> (N^2 synaptic weights, N input, non linearity parameter g). Additionally, synaptic weights are random.

It seems therefore impossible to characterize the dynamics of (3) in full generality.

(Dynamic) mean field method provides a way to characterize the "average" dynamics of (3) (averaged over J_{ij}'s) in the limit $N \to +\infty$.

Before developping this theory let us see what a <u>naive mean-field</u> theory would give. We denote $< >$ the average with respect to white noise ξ, and $[\ ]$ the average with respect to J_{ij}s. Note that $< >$ corresponds to averaging over a stochastic process, ξ, ~~where~~ where alea is renewed at each time; whereas $[\ ]$ corresponds to averaging over a "frozen" or "<u>quenched</u>" alea, once J_{ij}s are not renewed in time. An alternative formulation of the model would consider a version where J_{ij}s are renewed at each time too. The corresponding mean-field theory ("<u>annealed</u>") is quite easier to obtain.

Taking the average of (3) gives:

$$\frac{d\,[<V_i>]}{dt} = -[<V_i>] + \frac{1}{\sqrt{N}}\sum_{j=1}^{N}\left[W_{ij} < f_g(V_j(t))>\right] + \theta_i.$$

To simplify notation let us write $\mu_i \equiv [<V_i>]$.

To simplify further those equations let us make two simplifying assumptions (to be justified at the end of this chapter)

(i) <u>$V_j(t)$ is independent of W_{ij}.</u>

(ii) <u>Non linearity f_g and average commute.</u>

We shall discuss later on these two simplifications. For the moment let us proceed pragmatically. We obtain

$$\boxed{\frac{d\mu_i}{dt} = -\mu_i + \theta_i,} \qquad (4)$$

where we have used $[W_{ij}] = 0$. Note that this procedure easily extends to the general case of several neurons populations where the average value of J_{ij} connected j from population β to i (population α) is $\frac{\overline{J}_{\alpha\beta}}{N_\beta}$. We would obtain ($i \in$ pop α, $j \in$ pop β):

$$\frac{d}{dt}[\langle V_i \rangle] = -[\langle V_i \rangle] + \sum_{j=1}^{N_\beta} [J_{ij} \langle f_g(V_j) \rangle] + \Theta_i$$

$$\Rightarrow \quad \frac{d V_i}{dt} = -V_i + \frac{\overline{J}_{\alpha\beta}}{N_\beta} \sum_{j=1}^{N_\beta} f_g(V_j) + \Theta_i.$$

Let us define $\Theta_\alpha = \frac{1}{N_\alpha} \sum_{i=1}^{N_\alpha} \Theta_i$ the average stimulus population α. If we set $V_\alpha = \frac{1}{N_\alpha} \sum_{i=1}^{N_\alpha} V_i$ the average membrane potential in population α we have:

$$\frac{d V_\alpha}{dt} = -V_\alpha + \frac{\overline{J}_{\alpha\beta}}{N_\beta} \sum_{j=1}^{N_\beta} f_g(V_j) + \Theta_\alpha$$

Using once more the assumption of commutation between (empirical) average and non linearity, we finally obtain:

$$\boxed{\frac{d V_\alpha}{dt} = -V_\alpha + \overline{J}_{\alpha\beta} f_g(V_\beta) + \Theta_\alpha.} \qquad (5)$$

These phenomenological equations mimic e.g. the interaction between neurons populations in brain area such as cortical columns. A prominent example is Jansen-Rit equations which are used to mimic the effect of a thalamic input ($\Theta_\alpha \equiv \Theta_\alpha(t)$) on a set of neurons populations observed in cortical columns.

Surprisingly (taking into account the many questionable assumptions we made) this model gives fairly good results compared to experiments, for example in epilepsy. We shall comment later on reasons for this success on mathematical grounds. For the moment let us return to the simple case of eq. (9).

This equation is easy to integrate. It gives:

$$\mu_i(t) = \theta_i\left(1-e^{-t}\right) + \mu_i(0)\,e^{-t}, \qquad (6)$$

so that asymptotically $\mu_i(t) \to \theta_i$

The average $[\langle V_i \rangle] \equiv \mu_i$ converges to a <u>constant</u>, the stimulus value. This equation does not tell us that much on dynamics of (3) that one would expect quite more complex. Especially, since we take only the average of V_i, we have no information about <u>fluctuations</u> around the mean. So, a refinement of naive mean-field equations would consist of computing <u>higher order moments</u> of V_i. For this let us integrate (3). We have,

$$\frac{d}{dt}\left[e^t V_i\right] = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} W_{ij}\, f(V_j(t))\, e^t + \theta_i e^t + \sigma\, \xi_i(t)\, e^t$$

$$V_i(t)e^t - V_i(0) = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} W_{ij} \int_0^t f(V_j(s))\, e^s\, ds + \int_0^s \theta_i e^s\, ds + \sigma \int_0^s \xi_i(s)\, e^s\, ds$$

(we have left θ_i inside the integral to emphasize that this computation extends easily to a <u>time-dependent stimulus</u>). Finally

$$\boxed{V_i(t) = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} W_{ij} \int_0^t f(V_j(s))\, e^{s-t}\, ds + \theta_i\left(1-e^{-t}\right) + \sigma \int_0^s \xi_i(s)\, e^{s-t}\, ds + V_i(0)\, e^{-t}.} \qquad (7)$$

One sees that $V_i(t)$ depends, in a complex way, of the network history (membrane potentials $V_j(0)$) from the initial time to time t. It also depends on all W_{ij}'s. Let us now motivate assumption (i): since $V_i(t)$ depends on all W_{ij}'s, since there are N variables W_{ij}, and since their amplitude tends to 0 as $N \to +\infty$ we can make the assumption that, as $N \to +\infty$, $V_i(t)$ becomes independent of any given W_{ij}. This assumption is, in its spirit, very similar to Boltzmann's molecular chaos hypothesis. It has been introduced in the field of neural networks by Amari in 1972 under the name "local chaos hypothesis". Actually, Amari was also assuming the independence between V_i and V_j, $i \neq j$. Let us start without making this additional assumption. Using (i) we obtain:

$$[\langle V_i(t)\rangle] = \Theta_i(1-e^{-t}) + [\langle V_i(0)\rangle] e^{-t}$$

which is (6). Let us now compute the second order moments. For simplicity, let us assume $\Theta_i = 0$ and $V_i(0) = 0$ so that $[\langle V_i(t)\rangle] = \mu_i(t)$. (The following computation easily extends). We have:

$$[\langle V_i(t) V_k(t')\rangle] =$$

$$\frac{1}{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\Big[W_{ij}W_{kl}\Big\langle \int_0^t\int_0^{t'} f_g(V_j(0)) f_g(V_l(s')) e^{s-t} e^{s'-t'} ds\, ds'\Big\rangle\Big]$$

$$+\sigma^2\int_0^t\int_0^{t'} \underbrace{\langle \xi_i(s)\xi_j(s')\rangle}_{\delta_{ij}\,\delta(s-s')} e^{s-t} e^{s'-t'} ds\, ds'$$

$$+\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\left[W_{ij}\int_0^t\int_0^{t'}\langle f_g(V_j(s))\,\xi_k(s')\rangle e^{s-t}e^{s'-t'}\,ds\,ds'\right]$$

$$+\frac{1}{\sqrt{N}}\sum_{l=1}^{N}\left[W_{kl}\int_0^t\int_0^{t'}\langle f_g(V_k(s))\,\xi_i(s')\rangle e^{s-t}e^{s'-t'}\,ds\,ds'\right].$$

Using assumption (i), this gives:

$$[\langle V_i(t)V_k(t')\rangle] =$$

$$\frac{1}{N}\sum_{j,l=1}^{N}[W_{ij}W_{kl}]\int_0^t\int_0^{t'}[\langle f_g(V_j(s))f_g(V_l(s'))\rangle]e^{s-t}e^{s'-t'}\,ds\,ds'$$

$$+\delta_{ik}\sigma^2\int_0^{t\wedge t'}e^{2s-t-t'}\,ds,$$

where $t\wedge t'=\min(t,t')$. For simplicity and without loss of generality let us assume $t\le t'$ so that:

$$\boxed{\begin{aligned}&[\langle V_i(t)V_k(t')\rangle] = \\ &\frac{1}{N}\sum_{j,l=1}^{N}[W_{ij}W_{kl}]\int_0^t\int_0^{t'}[\langle f_g(V_j(s))f_g(V_l(s'))\rangle]e^{s-t}e^{s'-t'}\,ds\,ds' \\ &+\delta_{ik}\frac{\sigma^2}{2}\left[e^{t-t'}-e^{-t-t'}\right]\end{aligned}} \quad (8)$$

Higher order moments can be computed likewise.

Let us now make several remarks.

1) To compute each of these moments one has to know the law of trajectories $\{V(s)\}_{s=0}^{t}$ under the joint law $[\langle\ \rangle]$. This law is precisely defined via its moments. So, in principle one has to solve an infinite number of equations

corresponding to a hierarchy of moments, similarly as BBGKY hierarchy in hydrodynamics. However, this hierarchy can be cut at some order. This is in particular the case if V_i is Gaussian under the joint law [< >].

Looking at (2), one may in particular expect the sum to converge in law to a Gaussian process provided specific conditions, called Lindenberg conditions, are fulfilled. These conditions do not constrain W_{ij}s to be Gaussian. They ensure that central limit Theorem holds for (2).

We shall not state these conditions here (as we do not know anyway how to prove them). Instead, we shall work under the following hypothesis, that will be further commented later on.

Gaussian Hypothesis:

V converges in law to a Gaussian process as $N \to +\infty$, with mean $\mu = \{\mu_i(t)\}_{\substack{i=1 \ldots N \\ t=0 \ldots +\infty}}$ and covariance $\Delta(t,t')$ where $\Delta_{ik}(t,t') = [< V_i(t) V_k(t') >]$

Under this assumption one can truncate the infinite moment hierarchy to second order.

Gaussian hypothesis allows us to write the asymptotic mean-field limit of (3). Indeed, from the same hypothesis we may conjecture that the limit of (3) is an infinite dimensional equation of the form

$$\boxed{\frac{dV}{dt} = -V(t) + \eta(t) + \sigma\,\xi(t),} \qquad (9)$$

where V is the infinite dimensional process $(V_i)_{i\in\mathbb{N}}$ (one can also consider the bi-infinite case $(V_i)_{i\in\mathbb{Z}}$);

$\eta(t)$ is a <u>Gaussian process</u>, weak-limit of the synaptic field G defined in eq. (2). It has a mean 0 (when $\theta_i = 0$) and covariance given by the limit of:

$$E\left[G_i(t)\,G_k(t')\right] = \sum_{j,l=1}^{N} E\left[J_{ij} J_{kl}\right] \left[\langle f_g(V_j(t)) f_g(V_l(t')) \rangle \right.$$

as $N \to +\infty$.

2) <u>Independence of J_{ij}s:</u> Up to now we haven't used the assumption of independence of J_{ij}'s (W_{ij}'s), and actually our computation extend to that case. However, to illustrate the dynamic mean-field approach and specially the difficulty to solve the resulting equations, even in the simplest case, we shall restrict here to the case where J_{ij}s are <u>independent</u>. Then:

$$E\left[G_i(t) G_k(t')\right] = \frac{J^2}{N} \sum_{j,l=1}^{N} \delta_{ik}\,\delta_{jl} \left[\langle f_g(V_j(t)) f_g(V_l(t')) \rangle \right]$$

$$= \frac{J^2}{N} \sum_{j=1}^{N} \left[\langle f_g(V_j(t)) f_g(V_j(t')) \rangle \right] \delta_{ik}$$

So, that, taking the limit $N \to +\infty$:

$$[\langle \eta_i(t)\,\eta_k(t')\rangle] = \lim_{N\to+\infty} \frac{J^2}{N}\sum_{j=1}^{N} [\langle f(V_j(t))\,f(V_j(t'))\rangle]\,\delta_{ik}$$

We can further simplify this expression when θ_is are constant $\theta_i = \theta$ (here $\theta_i = 0$). In this case, indeed all V_j s have the same distribution and we obtain

$$[\langle \eta_i(t)\,\eta_k(t')\rangle] = J^2\,[\langle f(V_j(t))\,f(V_j(t'))\rangle]\,\delta_{ik}.$$

In this simplest example the infinite dimensional field η is the infinite product of one dimensional fields having the same distribution. So, from the assumption of independence between J_{ij}'s we transform the infinite dimensional problem into a one dimensional one. Therefore, adding correlations in synaptic weights lead to a drastic change of complexity.

Let us set:

$$C(t,t') = [\langle f(V(t))\,f(V(t'))\rangle]$$

so that

$$\boxed{[\langle \eta(t)\,\eta(t')\rangle] = J^2\,C(t,t'),} \qquad (10)$$

where η and V are now one dimensional, corresponding to a given neuron.

Now, V is Gaussian with mean zero ($\overline{O_i}=0$) and covariance $\Delta(t,t')$ given by the limit of (8), or by the integration of (9). Indeed, integrating (9) gives:

$$V(t) = \int_{t_0}^{t} \eta(s)\, e^{s-t}\, ds + \sigma \int_{t_0}^{t} \xi(s)\, e^{s-t}\, ds + V(t_0)\, e^{-(t-t_0)},$$

where we have taken the initial time at t_0. The advantage of this choice is that we can remove the dependence on initial condition by taking $t_0 \to -\infty$. Then:

$$V(t) = \int_{-\infty}^{t} \eta(s)\, e^{s-t}\, ds + \sigma \int_{-\infty}^{t} \xi(s)\, e^{s-t}\, ds, \qquad (11)$$

so that:

$$\Delta(t,t') \equiv [\langle V(t)V(t')\rangle] =$$

$$\int_{-\infty}^{t}\int_{-\infty}^{t'} [\langle \eta(s)\eta(s')\rangle]\, e^{s-t}\, e^{s'-t'}\, ds\, ds' + \sigma^2 \int_{-\infty}^{t}\int_{-\infty}^{t'} \langle \xi(s)\xi(s')\rangle\, ds\, ds'$$

$$+ \sigma \int_{-\infty}^{t}\int_{-\infty}^{t'} \left([\langle \eta(s)\xi(s')\rangle] + [\langle \eta(s')\xi(s)\rangle] \right) e^{s-t}\, e^{s'-t'}\, ds\, ds'.$$

This is the limit $N \to +\infty$ of (8) and, by comparison, one observes that the third term cancels. We finally obtain:

$$\boxed{\Delta(t,t') = J^2 \int_{-\infty}^{t}\int_{-\infty}^{t'} C(s,s')\, e^{s-t}\, e^{s'-t'}\, ds\, ds' + \frac{\sigma^2}{2}\, e^{t-t'} \qquad (12) \qquad (t \le t')}$$

Let us summarize and comment what we have obtained:

(Under the many assumptions made)

1) The solution of (3), averaged over the stochastic noise and the synaptic weight randomness is "equivalent", in the limit $N \to +\infty$, to the solution of:

$$\boxed{\frac{dV}{dt} = -V + \eta + \sigma\xi,} \quad (\text{eq.}(9))$$

where

* V is one dimensional in the case where synaptic weights are independent and centered; it is infinite dimensional if weights are correlated;

* η is a Gaussian process, centered with covariance

$$\boxed{J^2 C(t,t') = J^2 [< f(V(t)) f(V(t')) >]} \quad (\text{eq.}(10))$$

* The covariance of V, Gaussian centered, is

$$(\text{eq.}(12)) \quad \boxed{\Delta(t,t') = J^2 \int_{-\infty}^{t}\int_{-\infty}^{t'} e^{s-t} e^{s'-t'} C(s,s')\, ds\, ds' + \frac{\sigma^2}{2} e^{-|t-t'|}}$$

2) Since V is Gaussian, the quantity $C(t,t')$ can be easily computed. It gives rise to a self-consistent equation, to be written and solved below.

3) Looking at (9) one observes that the evolution of V

is driven by two terms: the white noise and the local synaptic field.). This last term is the result of an average over noise and synaptic weight randomness (quenched randomness) as easily seen from (10).

Although it looks quite simple the analysis and interpretation of this local field is in fact quite complex. To our best knowledge it has been introduced and analysed for the first time by Sompolinsky and Zippelius, in 1981, for the study of spin glasses.

We analyse its properties in the next section.

(4) If we compare with naive mean-field theory we have introduced, as expected, an effective dynamical equation characterizing, not only the average of V, but also its fluctuations. Compared with chapter I (Curie-Weiss model) where the order parameter was magnetization and where the phase transition was characterized by a static quantity, we have here to handle a dynamical equation, where order parameters remain to be identified.

4) Solutions of Dynamic Mean Field Equations – Stationarity assumption.

for $\sigma = 0$

In order to solve these equations we make the assumption that the process described by (9) is stationary. As a consequence: $\Delta(t,t') = \Delta(t'-t)$, $C(t,t') = C(t-t')$ and:

$$\Delta(t'-t) = J^2 \int_{s=-\infty}^{t} \int_{s'=-\infty}^{t'} e^{s-t} e^{s'-t'} C(s'-s)\, ds\, ds' + \frac{\sigma^2}{2} e^{-(t'-t)}$$

$(t'=0, t=\tau)$

$$\Rightarrow \Delta(\tau) = J^2 \int_{s'=-\infty}^{0} \int_{s=-\infty}^{\tau} e^{s'} C(s'-s) e^{s-\tau}\, ds\, ds' + \frac{\sigma^2}{2} e^{-|\tau|}$$

Introducing the convolution product

$$(f*g)(x) = \int_{-\infty}^{+\infty} f(y)\, g(x-y)\, dy,$$

the notation $\tilde{f}(x) = f(-x)$, and the function $\gamma(\tau) = \begin{cases} e^{-\tau}, & \text{if } \tau \geq 0, \\ 0, & \text{otherwise}, \end{cases}$ we see that

$$(\gamma * C)(u) = \int_{0}^{+\infty} e^{-v}\, C(u-v)\, dv;$$

$$(\widetilde{\gamma * C})(u) = \int_{0}^{+\infty} e^{-v}\, C(-u-v)\, dv;$$

$$\gamma * (\widetilde{\gamma * C})(-\tau) = \int_{-\infty}^{+\infty} \gamma(-\tau-u)\, (\widetilde{\gamma * C})(u)\, du$$

$$= \int_{u=-\infty} e^{(\ldots)} \left(\int_{v=0} e^{-v} C(-u-v)\, dv \right) du$$

$$\overset{t=-\tau-u}{=} \int_{t=0}^{+\infty} e^{-t} \int_{v=0}^{+\infty} e^{-v} C(\tau+t-v)\, dv\, du.$$

Setting $s'=-t$, $s=\tau-v$ we finally obtain:

$$\gamma * (\widetilde{\gamma * C})(-\tau) \equiv \left(\widetilde{\gamma * (\widetilde{\gamma * C})}\right)[\tau] =$$

$$\int_{s'=-\infty}^{0} \int_{s=-\infty}^{\tau} e^{s'} C(s-s') e^{s-\tau}\, ds\, ds'.$$

Thus $$\boxed{\Delta(\tau) = J^2 \left[\widetilde{\gamma * (\widetilde{\gamma * C})}\right][\tau] + \frac{\sigma^2}{2} e^{-\tau}.}$$

Denoting by $\hat{f}(\omega) = \int_{-\infty}^{+\infty} e^{-i\omega\tau} f(\tau)\, d\tau$ the Fourier transform of f we have.

$$\hat{\gamma}(\omega) = \frac{1}{1+i\omega} \quad ; \quad \hat{\tilde{\gamma}}(\omega) = \frac{1}{1-i\omega},$$

$$\widehat{\gamma * (\widetilde{\gamma * C})} = \hat{\tilde{\gamma}}(\omega)\, \hat{\gamma}(\omega)\, \hat{C}(\omega) = \frac{\hat{C}(\omega)}{1+\omega^2}.$$

Therefore:

$$\hat{\Delta}(\omega) = J^2 \frac{\hat{C}(\omega)}{1+\omega^2} + \frac{\sigma^2}{1+\omega^2} \quad \text{or} \quad J^2 \hat{C}(\omega) = \hat{\Delta}(\omega) + \omega^2 \hat{\Delta}(\omega) - \sigma^2$$

Since $\omega^2 \hat{f}(\omega) = -\widehat{\left(\frac{d^2 f}{dt^2}\right)}$ we finally obtain, taking the inverse Fourier transform:

$$\boxed{\Delta - \frac{d^2\Delta}{d\tau^2} = J^2 c(\tau) + \sigma^2 e^{-|\tau|}} \quad (13)$$

Let us now show that $c(\tau)$ is a function of $\underline{\Delta(\tau)}$ and $\underline{\Delta(0)}$
Indeed:

$$c(\tau) = [\langle f(V(\tau)) f(V(0)) \rangle]$$

where $(V(0), V(\tau))$ is Gaussian centered with covariance matrix:

$$Q = \begin{bmatrix} \Delta(0) & \Delta(\tau) \\ \Delta(\tau) & \Delta(0) \end{bmatrix}; \quad (14)$$

Therefore:

$$c(\tau) = \frac{1}{2\pi \det Q^{1/2}} \int_{\mathbb{R}^2} e^{-\frac{1}{2}\tilde{X} Q^{-1} X} f(x) f(y)\, dx\, dy$$

$X = \begin{pmatrix} x \\ y \end{pmatrix} \Rightarrow$

$$\boxed{c(\tau) = \frac{1}{2\pi\sqrt{\Delta^2(0) - \Delta^2(\tau)}} \int_{\mathbb{R}^2} e^{-\frac{1}{2}\left[\frac{2\Delta(0)(x^2+y^2) - 2xy\Delta(\tau)}{\sqrt{\Delta^2(0)-\Delta^2(\tau)}}\right]} f(x) f(y)\, dy} \quad (15)$$

... convenient to use two other representations. First, introducing the scalar product

$$\langle f, g \rangle = \int_{\mathbb{R}} f(x) g(x)\, dx,$$

and the functions $\Psi(x,y) = f_g(x) f_g(y)$,

$$\eta(x,y) = \frac{1}{2\pi \det Q^{1/2}} e^{-1/2\, \tilde{x}\, Q^{-1} x},$$ we have:

$$C(\tau) = \langle \Psi, \eta \rangle.$$

Defining the spatial Fourier transform $\hat{f}(k) = \int_{-\infty}^{+\infty} e^{-ikx} f(x)\, dx$ we obtain from Parseval's theorem:

$$C(\tau) = \langle \hat{\Psi}, \hat{\eta} \rangle,$$

where $\hat{\Psi}(k,k') = \hat{f}_g(k) \hat{f}_g(k')$ and

$$\hat{\eta}(k,k') = \frac{1}{2\pi} e^{-\frac{1}{2}[\Delta(0)(k^2+k'^2) + 2\Delta(\tau) k k']}$$

Therefore:

$$\boxed{C(\tau) = \int \hat{f}_g(k) \hat{f}_g(k') e^{-\left\{\frac{\Delta(0)}{2}(k^2+k'^2) + k k' \Delta(\tau)\right\}} \frac{dk}{2\pi} \frac{dk'}{2\pi}} \quad (16)$$

From (15) one can also show (see added notes) that:

$$C(\tau) = \int_{-\infty}^{+\infty} \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \left[\int_{-\infty}^{+\infty} \frac{dx}{\sqrt{2\pi}} e^{-x^2/2} f_g\left(\sqrt{\Delta(0) - |\Delta(\tau)|}\, x + \sqrt{|\Delta(\tau)|}\, z \right) \right]^2$$

(17).

What's the use of these formula? We see that, as announced $C(\tau)$ is in fact a function of $\Delta(0)$, $\Delta(\tau)$
$C(\tau) \equiv H(\Delta(0), \Delta(\tau))$. $\Delta(0)$ and $\Delta(\tau)$ play therefore the role of parameters (in fact <u>order parameters</u>). We shall come back to this point later on.

The important issue coming out from eq. (15), (16), (17) is that they make easy a power expansion of $H(\Delta(0), \Delta(\tau))$ in powers of $\Delta(\tau)$ (Why not $\Delta(0)$? we shall explain this later). Indeed, from (16) the derivation ~~of~~ H with respect to $\Delta(\tau)$ generates a factor $-kk' = ik\,ik'$. Since $ik\,\hat{f}(k)\,ik\,\hat{f}(k')$ is the Fourier transform of $\frac{df_g}{dx} = g\,f_g'(x)$, the identification (17) gives

$$\frac{d\,C}{d\Delta(\tau)} = g^2 \int_{-\infty}^{+\infty} \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \left[\int_{-\infty}^{+\infty} \frac{dx}{\sqrt{2\pi}} e^{-x^2/2} f_g'\left(\sqrt{\Delta(0) - |\Delta(\tau)|}\, x + \sqrt{|\Delta(\tau)|}\, z \right) \right]^2 \qquad (18)$$

From the same reasoning

$$\boxed{\frac{d^n C}{d\Delta^n(\tau)} = g^{2n} \int_{-\infty}^{+\infty} \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \left[\int_{-\infty}^{+\infty} \frac{dx}{\sqrt{2\pi}} e^{-x^2/2} f_g^{(n)}\left(\sqrt{\Delta(0) - |\Delta(\tau)|}\, x + \sqrt{|\Delta(\tau)|}\, z \right) \right]^2}$$

Therefore, we have a <u>Taylor expansion</u> of C in powers of $\Delta(\tau)$.

$$\boxed{C(\tau) = \sum_{n=0}^{+\infty} \frac{\Delta(\tau)^n}{n!} \frac{d^n C}{d\Delta^n(\tau)}} \quad , \qquad (20)$$

where the coefficients can be explicitly computed (at least numerically)!

Let us now return to the Newtonian equation (13). It can be qualitatively analysed by setting $\boxed{(\text{for } \sigma = 0)}$

$$\boxed{\Delta - J^2 C(\tau) = -\frac{\partial V}{\partial \Delta(\tau)}} \quad (21)$$

Indeed (13) becomes:

With the constraints (correlation)

$$\boxed{\begin{aligned} \Delta(\tau) &= \Delta(-\tau) \\ |\Delta(\tau)| &\leq \Delta(0) \end{aligned}} \quad (\mathcal{C})$$

$$\boxed{\ddot{\Delta} = -\frac{\partial V}{\partial \Delta(\tau)}} \quad , \qquad (22)$$

the equation of a particle motion in a potential well, where from (20), (21) (setting $\Delta(\tau) \equiv \Delta$ to alleviate notations)

$$\boxed{V(\Delta) = K - \frac{1}{2}\Delta^2 + J^2 \sum_{n=1}^{+\infty} \frac{\Delta^n}{n!} V_n} \quad , \qquad (23)$$

with.

$$V_n = g^{2n-2} J^2 \int_{-\infty}^{+\infty} \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \left[\int_{-\infty}^{+\infty} \frac{dx}{\sqrt{2\pi}} e^{-x^2/2} f_g^{(n-1)}\left(\sqrt{\Delta(0) - |\Delta(0)|}\, x + \sqrt{|\Delta(0)|}\, z \right) \right]^2 \qquad (24)$$

Note that we have taken the derivatives at $\underline{\Delta(\tau) = \Delta(0)}$

Let us now explain why.

(The constant K can be chosen so that $V(0) = 0$.)

5) Bifurcations of dynamic-mean field equations.

The quantity $\Delta(0)$ corresponds to the initial value of the correlation function of a neuron membrane potential, in the stationary regime. As a consequence this is not a free parameter: its value is fixed by dynamics and by control parameters (here g, J, more generally Θ_i - -)

It plays the same role as the magnetization in Curie-Weiss model and is an order parameter.

Now, $\Delta(\tau)$ characterizes the evolution the correlation function as τ increases, starting from $\Delta(0)$. This evolution is controlled by Newton equation, where $V(\Delta)$ is precisely given by a Taylor expansion at $\Delta_0 = \Delta(0)$.

The shape of $V(\Delta)$ controls dynamics and bifurcations. In the present example it is controlled by the non linearity parameter g. Let us investigate a few examples.

5(i) $f_g(x) = \tanh(gx)$. This is the case studied by Sompolinsky and coworkers in 1988 (Sompolinsky, Crisanti, Sommers, PRL, 1988).

In this case, $f_g(x)$ has the symmetry $x \to -x$, so that all odd terms in (23) vanish.

The lowest order term, Δ^2 has a coefficient

$-1 + g^2 J^2 \left[f_g'^{2} \right]_{\Delta_0}$, where we note, for simplicity

$$[\Psi]_{\Delta_0} = \int_{-\infty}^{+\infty} \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \left[\int_{-\infty}^{+\infty} \frac{dx}{\sqrt{2\pi}} e^{-x^2/2} \varphi\left(\sqrt{\Delta_0 - |\Delta|}\, x + \sqrt{|\Delta|}\, z\right) \right]^2_{\Delta = \Delta_0}$$

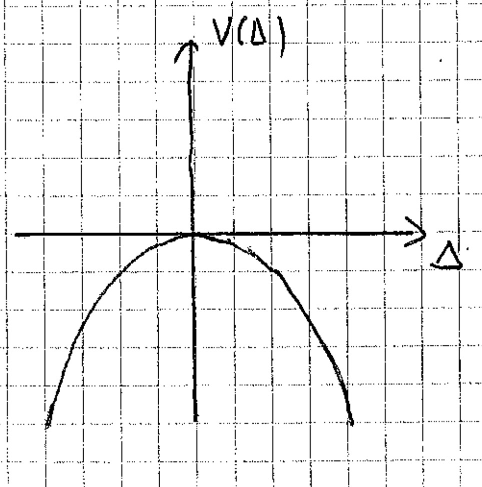
and, once here we can take $\Delta_0 > 0$ without loss of generality:

$$\boxed{[\Psi]_{\Delta_0} = \int_{-\infty}^{+\infty} \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \varphi^2\left(z\sqrt{\Delta_0}\right).}$$

The shape (convexity) of $V(\Delta)$ is controlled by the lowest order term. Thus,

* if $g^2 J^2 [f_g']_{\Delta_0} < 1$ V is concave at the origin.

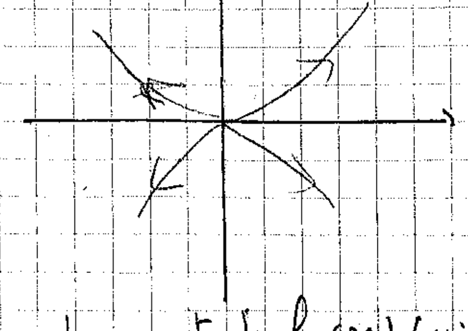
Now, once $\Delta(t)$ is a correlation function it has to obey the constraints (8). especially, $|\Delta(\tau)| \leq \Delta_0$ so that the only possible solution is



$$\boxed{\Delta(\tau) = \Delta_0 = 0}. \quad (25)$$

Since $f_g'(x) = 1 - th^2(gx)$, the convexity of V is controlled by the condition:

(26) $\boxed{gJ = 1}$ (change of convexity)



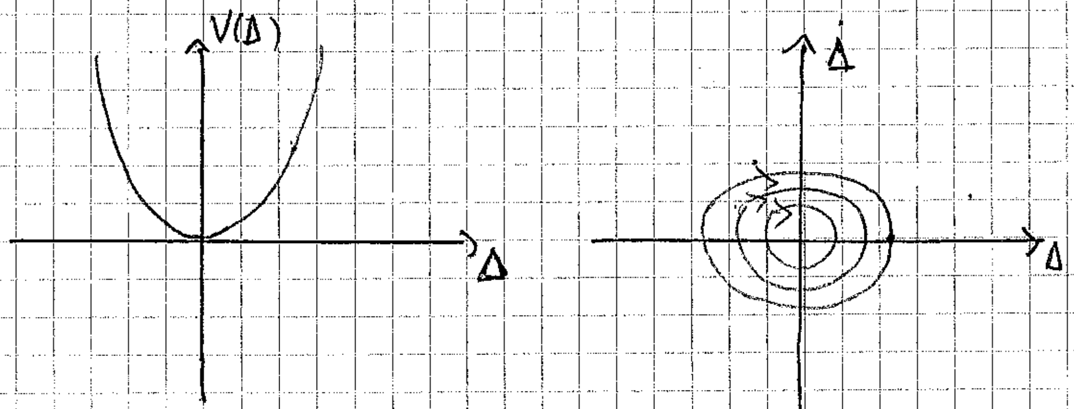
The process solution of (9) is a Gaussian centered, with zero covariance (if $\sigma = 0$).

phase portrait of (22) (without constraint 8)

This is therefore a "process" with almost surely constant trajectories, all equal to 0. We interpret them in terms of the finite dimensional system (3) in the next section.

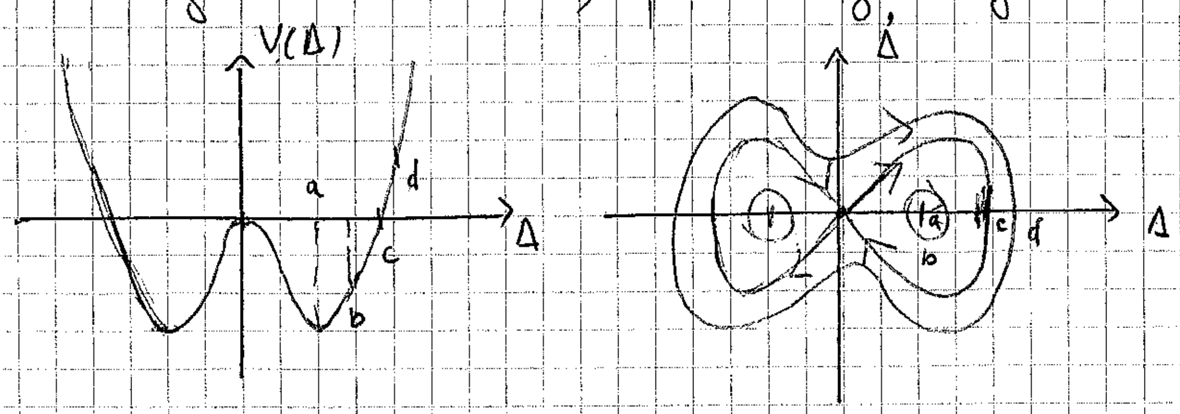
(In the presence of white noise, this picture does not hold anymore, but we have anyway to reconsider the dynamical equation (22) for the covariance. This is done below).

* if $g^2 J^2 > 1$, V becomes convex at the origin. Depending on the sign of higher order term we have two possibilities



phase portrait

$\Delta(\tau)$ is periodic with a maximum Δ_0. There are infinitely many such solutions, parametrized by Δ_0.



There are 4 possible characteristic shapes

a) Fixed point with $\Delta_0 > 0$

b) Periodic orbit centered in a.

c) homoclinic orbit

d) Periodic orbit centered in 0.

These 4 cases correspond in terms of the stochastic mean-field process (9) to:

a) Constant covariance $\Delta_0 = \Delta(\tau)$.

$\Rightarrow$ Process with almost surely constant trajectories.

b), d) Periodic covariance

$\Rightarrow$ Process with 0 mean and periodic variation of auto-correlation.

c) $\Delta_0 > 0$ and $\Delta(\tau) \to 0$ as $\tau \to +\infty$.

This is interpreted by Sompolinsky et al as a signature of a mixing (chaotic) process, although it could be the signature of an Ornstein-Uhlenbeck process. Actually... this is both, as explained in the next section.

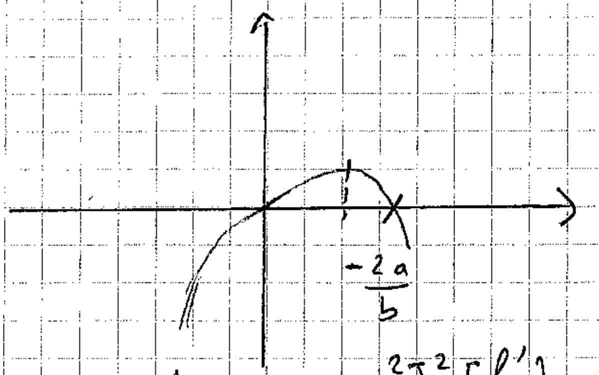
5.ii) $f_g(x)$ has not the symmetry $x \to -x$

In this (23) has generically all terms non vanishing, and the lowest order term in $V(\Delta)$ is $J^2[f_g]_{\Delta_0} \Delta$, so that, up to second order:

$$V(\Delta) \sim J^2 [f_g]_{\Delta_0} + \left(g^2 J^2 [f_g]_{\Delta_0} - 1\right) \frac{\Delta^2}{2} + \cdots$$

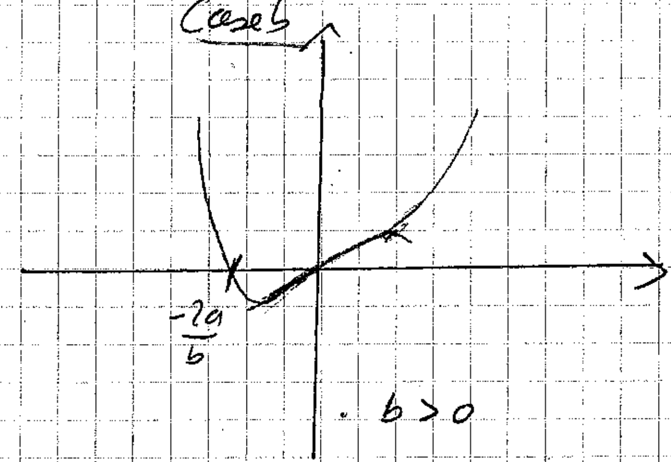
Taking $f_g \geq 0$ (this is a firing rate) the typical shape of V near the origin is:

Case a



$b<0 \Rightarrow g^2 J^2 [f_g']_{\Delta_0} < 1$

Case b



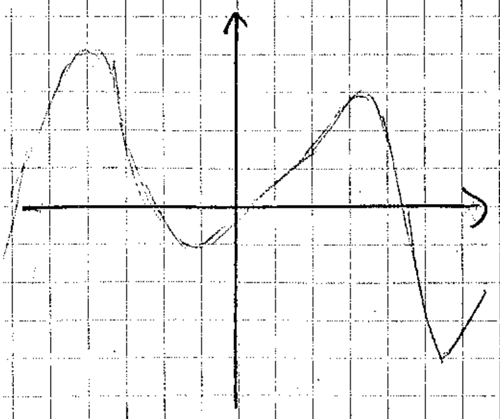
depending on the sign of $b = g^2 J^2 [f_g']_{\Delta_0} - 1$.

1) If $g^2 J^2 [f_g']_{\Delta_0} < 1$ there is only one admissible solution at $\Delta(t) = \Delta_0 = \dfrac{1 - g^2 J^2 [f_g']_{\Delta_0}}{J^2 [f_g]_{\Delta_0}}$.

This corresponds, again, to a process with almost surely constant trajectories.

2) If $g^2 J^2 [f_g'] > 1$ there are oscillating solutions (not symmetric with respect to $\Delta = 0$) around the minimum of the potential $\dfrac{1 - g^2 J^2 [f_g] \Delta_0}{J^2 [f_g]_{\Delta_0}}$

Higher order terms complexify the shape of V. Especially



the terms of order 3, 4 may be responsible for the existence of multiple solutions. They appear, as g increases, by saddle-node bifurcations.

Symmetry breaking is also induced by adding a stimulus $\Theta_i = \Theta \neq 0$. (If the stimulus is not spatially homogeneous there is not one equation but possibly infinitely many (as many as distinct Θ_i's)).

5(iii) Role of noise

When $\sigma \neq 0$, there is an additional term to Newton's equation (22):

$$\ddot{\Delta} = -\frac{\partial V}{\partial \Delta} - \sigma e^{-|z|},$$

This corresponds to an additional, time dependent force. It can interpreted, still in the picture of a particle in a potential well, as a time decaying gravitional field, oriented upward (negative sign).

6) Going beyond

To finally arrive ~~to~~ eq. (22) and its solutions, we have made a lot of approximations / hypotheses that we want now to discuss. We also want to discuss their interpretation in terms of the initial system, the Amari - Cohen - Grossberg model (1).

6.1) Avoiding the local chaos hypothesis.

To obtain eq. (9) (dynamic mean - field equation), we have made the assumption of independence between V_j and J_{ij}. Can we obtain ~~the~~ equation (9) without making this hypothesis? The answer is "yes" and there are in fact two ways of obtaining it.

(1) Generating functional

The first method, based on a similar approach as the computation of the partition function in chap I, consists of computing a functional ("Feynman integral") generating the moment of V, averaged over [ ] and < >. This method has been introduced by Sompolinsky and Zippelius for spin glasses, in 1981, on the basis of previous work by De Dominicis (78) and Martin - Siggia - Rose (75).

This functional, with a form similar (although more complex) as the Curie - Weiss partition function, is computed, in the limit $N \to +\infty$, using a steepest descent method. When J_{ij}'s are independent, this functional decomposes as an infinite product of generating functionals for one dimensional processes precisely corresponding to (9). One obtains therefore the Dynamic Mean Field Equation without using the local chaos hypothesis. But the reason why this method works is precisely what allowed us to "risk" ~~the~~ use the local chaos hypothesis: (i) fully-connected model; (ii) synaptic weights with an amplitude tending to 0 as $N \to +\infty$.

Although apparently more satisfying than local chaos this method has several drawbacks and caveats. ~~In~~ The course of computation several ad hoc hypotheses are introduced such as the introduction then cancelling of an additional field. Moreover, this method does not easily extend to the correlated case. This is certainly, to our best knowledge, no result have been published on DMFT for correlated weights using this technique.

(ii) Large deviations

In the seminal paper (95) Ben Arous and Guionnet have attempted to justify the Dynamic Mean Field approach of Sompolinsky-Zippelius for spin glasses. To the price of high level techniques they have obtained a rigorous derivation of these equations using large deviations. Later on, Moynot and Samuelides (2002) have applied this method to a discrete time version of Amari-Wilson-Cowan, formerly studied by Cessac et al (94, 95). Finally, more recently, Faugeras and MacLaurin (2014) have extended this approach to a correlated version of Cessac et al model.

Beyond the fact that this method is rigorous it has the advantage to extend to the correlated case, contrarily to functional generating method.

Additionally, the knowledge of the rate function could be useful for further interpretations of the dynamics of mean-field equations, and their correspondence with the finite size system.

6.2) Interpreting mean-field equations

6.2.1.) Finite size dynamical system.

The dynamics of (1) can be essentially explored by numerical simulations although a couple of mathematical results can be obtained. In the case studied here (independent synaptic weights, balanced case) the most prominent aspects of dynamics are :

(1) Transition to chaos by quasi periodicity as g increases

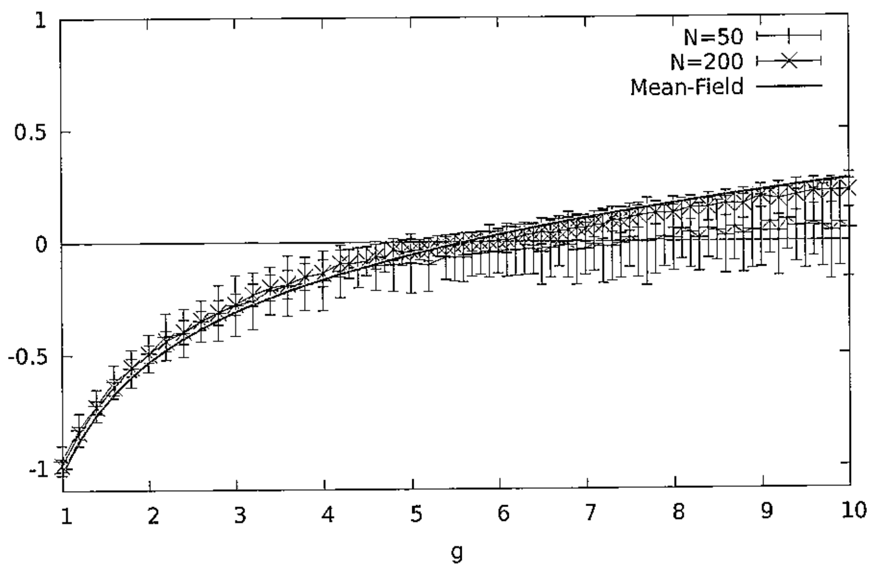
(2) Coexistence of several dynamical attractors in presence of a stimulus. The appearance or disappearance of attractors occurs by saddle-node bifurcations.

The location of attractors in the phase space, as well as the parameters (g, θ) value for bifurcation are random: they depend upon the realization of synaptic weights.

The transition to chaos when increasing g is sharper and sharper as N increases, suggesting a sharp transition fixed point - chaos in the limit $N \to +\infty$.

The characterization of dynamics can be completed by quantitative observations such as Lyapunov spectrum or Kolmogorov - Sinai entropy.

As an example we have drawn in the next figure the maximal Lyapunov exponent as g increases, with error bars corresponding to several realizations of synaptic weights.



6.2.2) Mean-field theory versus finite size system.

Transition to chaos corresponds to a wide variety of dynamical regimes from fixed to periodic, quasiperiodic, chaos. Quantitatively this transition is characterized by the Lyapunov spectrum and more specifically by the maximal Lyapunov exponent. At the transition it exhibits a plateau at 0, before growing to positive values when g increases (see previous figure). On the opposite, the Mean-Field Lyapunov exponent increases monotonously and therefore it does not exhibit this plateau. This is, in fact in perfect agreement with numerical observations! Indeed, the width of the plateau (the range of g values for which $\lambda = 0$) decreases and tends to 0 as $N \to +\infty$. Therefore, the mean-field Lyapunov exponent correctly predicts the asymptotic behaviour of the neural network (3) in the thermodynamic limit.

However the corollary of this statement, is that the mean-field theory does not describe correctly the finite size system: it misses many essential aspects of the dynamics, such as:

(i) Multiplicity of dynamical regimes and plateau of the Lyapunov exponent (as said above)

(ii) Resonances: when submitted to a periodic stimulus of weak amplitude, the model (3) exhibits a surprisingly rich structure of resonances, depending on J_{ij}'s. On the opposite mean-field theory predicts a rather poor resonance variety

(iii) Commutation of limits. This is not properly a caveat of mean-field theory but instead an open problem. The classical view of dynamical systems theory considers a finite-dimensional system, such as (3), and investigates the structure of its ω-limit set (union of attractors). When one performs numerical simulations with increasing N one considers the limits:

$$\lim_{N\to+\infty} \lim_{t\to+\infty}$$

in this order. Mean-field theory considers the opposite order. Although, numerical simulations shows a good agreement with mean-field theory and do not exhibit a clear problem of limit commutation, this aspect has, to the best of our knowledge, not been systematically explored, in particular for more general synaptic weights distribution (correlated). Note that, in spin glasses, these two limits do not commute at low temperature (this is known as the problem of "aging").

6.23) Back to naive mean field theory.

Let us now come back to the multi-populations model discussed above in the realm of naive mean-field theory. Using the wisdom acquired from this lecture we can now write the correct equations. Thus, let us consider p populations of neurons, with N_α neurons in population α. A synaptic weight J_{ij} ($j \in \beta$, $i \in \alpha$) has now a Gaussian distribution $\mathcal{N}\left(\frac{\bar{J}_{\alpha\beta}}{N}, \frac{J^2_{\alpha\beta}}{N}\right)$. We assume J_{ij}'s to be independent.

Equation (3) becomes:

$$\boxed{\frac{dV_i}{dt} = -\frac{V_i}{\tau_\alpha} + \sum_{\beta=1}^{P} \sum_{j=1}^{N_\beta} J_{ij} f_\beta(V_j(t)) + \theta_\alpha + \sigma_\alpha \xi_\alpha(t)} \quad (27)$$

where we have added a characteristic time constant τ_α and a sigmoid function f_β depending on the population. Here the stimulus and noise depend on the population as well. This model has been analyzed in Faugeras, Touboul, Cessac, 2009.

The dynamic mean-field approach leads to an effective equation per populations:

$$\boxed{\frac{dV_\alpha}{dt} = -\frac{V_\alpha}{\tau_\alpha} + \eta_\alpha(t) + \theta_\alpha + \sigma_\alpha \xi_\alpha(t)} \quad (28)$$

where η_α is a Gaussian field summarizing the effect action of pre-synaptic neurons on a neuron in population.

More precisely, we have:

$$\eta_\alpha(t) = \sum_{\beta=1}^{P} \eta_{\alpha\beta}(t),$$

where:

$$[\langle \eta_{\alpha\beta}(t)\rangle] = \overline{J}_{\alpha\beta}[\langle f_\beta(V_\beta(t))\rangle].$$

$$\mathrm{Cov}(\eta_{\alpha\beta}(t), \eta_{\gamma\delta}(0)) = \overline{J}^2_{\alpha\beta}[\langle f_\beta(V_\beta(t))f_\beta(V_\beta(0))\rangle]\,\delta_{\alpha\gamma}\delta_{\beta\delta}$$

(neurons are independent due to the assumption of independence between J_{ij} s). In particular,

$$\mathrm{Var}(\eta_{\alpha\beta}(t)) = \overline{J}^2_{\alpha\beta}[\langle f^2(V_\beta(t))\rangle]$$

For the mean, μ_α, of V_α, we have therefore:

$$\frac{d\mu_\alpha}{dt} = -\frac{\mu_\alpha}{\tau_\alpha} + \sum_{\beta=1}^{P}\overline{J}_{\alpha\beta}[\langle f_\beta(V_\beta(t))\rangle] + \theta_\alpha + \sigma_\alpha\xi_\alpha(t),$$

and since V_α is Gaussian with variance, say $v_\beta(t)$, we have:

$$\boxed{\frac{d\mu_\alpha}{dt} = -\frac{\mu_\alpha}{\tau_\alpha} + \sum_{\beta=1}^{P}\overline{J}_{\alpha\beta}\int_{-\infty}^{+\infty} f_\beta\left(\sqrt{v_\beta(t)}\,h + \mu_\beta(t)\right)\frac{e^{-h^2/2}}{\sqrt{2\pi}}\,dh + \theta_\alpha + \xi_\alpha(t).} \quad (29)$$

In order to obtain the equation of the mean, we need to know the equation of the variance, and therefore to solve the full equation for the covariance of V_α s, as we did before, but here for a multi-population model. We find (see Faugeras et al, 2009)

$$\Delta_{\alpha\beta}(t,s) \equiv \mathrm{Cov}\left[V_\alpha(t)\,V_\beta(s)\right] =$$

$$\delta_{\alpha\beta}\, e^{-(t+s)/\tau_\alpha}\Big[v_\alpha(0) + \frac{\tau_\alpha \sigma_\alpha^2}{2}\left(e^{2s/\tau_\alpha} - 1\right)$$

$$+ \sum_{\beta=1}^{P} J_{\alpha\beta}^2 \int_0^t \int_0^s e^{(u+v)/\tau_\alpha}\, C_\beta(u,v)\, du\, dv\Big],$$

$$C_\beta(t,s) = \left[\langle f_\beta(V_\beta(t))\, f_\beta(V_\beta(s))\rangle\right] =$$

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_\beta\left(h\frac{\sqrt{v_\beta(t)\,v_\beta(s) - \Delta_{\beta\beta}(t,s)}}{\sqrt{v_\beta(s)}} + h'\frac{\Delta_{\beta\beta}(t,s)}{\sqrt{v_\beta(s)}} + \mu_\beta(t)\right)$$

$$\times f_\beta\left(h'\sqrt{v_\beta(s)} + \mu_\beta(s)\right) \frac{e^{-h^2/2}}{\sqrt{2\pi}} \frac{e^{-h'^2/2}}{\sqrt{2\pi}}\, dh\, dh'$$

These equations have, to our best knowledge, not been solved analytically. Even their numerical simulation is cumbersome, since it requires an integration over the whole trajectory ($[0,t]$ of $C_\beta(t,s)$).

To conclude with the link with naive mean-field approach, one easily sees that, if $J_{\alpha\beta} = 0$, the coupling term between populations vanish. Moreover, as $t \to +\infty$ $\Delta_{\alpha\beta}(t,s) \to 0$ so that the variance $v_\beta(t) \equiv \Delta_{\beta\beta}(t,t) \to 0$. As a consequence, eq (29) reduces to:

$$\frac{d\psi_\alpha}{dt} = -\frac{\psi_\alpha}{\tau_\alpha} + \sum_{\beta=1}^{P} \bar{J}_{\alpha\beta} f_\beta(\psi_\beta(t)) + \theta_\alpha + \xi_\alpha(t),$$

which is precisely equation (5), obtained in the realm of naive mean-field approximation.

Therefore, we see that the naive mean-field approach, leading in particular to Jansen-Rit equa[tions] and obtained under a questionable assumption of commutation between empirical average and non linearity are in fact correct, if synaptic weights in a population are replaced by the average value $\bar{J}_{\alpha\beta}$, i.e. fluctuations of synaptic weights are neglected.

Considering these fluctuations leads to considerable complexification of equations and analysis.

6.3) Open questions

6.3.1) Getting closer to neuroscience

In my opinion, to become really relevant to neuroscience mean-field methods have to considerably extend their scope, essentially limited by the techniques used to derive the mean-field equations, and, mostly to solve them. Let us list a set of possible and necessary extensions.

(i) Dealing with several populations.

Using local chaos hypothesis, it is easy to obtain the Dynamic Mean Field Equations for several populations in AWC model (see Faugeras, Touboul, Cessac, 2008). But, writing those equations is just the beginning of the story. To our best knowledge, there is neither mathematical nor analytical result on mean-field dynamics in this case; only numerical simulations. Note that the analysis made in this lecture, based on the analogy between equation of covariance and Newton's equation in a potential well does not apply.

(ii) Correlated weights.

Here too local chaos hypothesis provides a way to derive the dynamic mean-field equations, confirmed by the recent rigorous paper by Faugeras and MacLaurin (2014). The analysis of these equations is under investigation (Cessac et al in progress)

iii) Synaptic plasticity

Synaptic weights correlations are generated by dynamical processes such as synaptic or intrinsic plasticity. When submitted to these biophysical inspired mechanism, model (3) exhibit astonishing properties such as learning patterns, with optimal performances at the edge of chaos. No mean-field of plasticity exists yet.

iv) Spiking models.

Model (3) is a firing rate model, where spiking activity is averaged over a time window (typically ~ 100 ms). For many examples, it is necessary to have a description of neurons dynamics at the time scale of a spike (~ 1 ms). Model examples are Hodgkin-Huxley's model, FitzHugh-Nagumo, Morris-Lecar, Integrate and Fire, and so on. To our best knowledge, no satisfactory mean-field theory exist yet in this context.

v) Spatially extended models

Since mean-field methods allow an integration over spatial scales, they could be used to construct rigorously equations for neural masses and neural fields. It is not clear that these equations – where, as in model (3), fluctuations about the mean play a role as important as the mean itself – will have the same structure as the neural field equations currently used.

6.3.2) Extensions to other domains.

The dynamic mean field approach developped here does not use that much the neuro inspired structure of the model. It could therefore be applied in other contexts such as population dynamics, economics (?). The main constraint is the form of the coupling field which appears as a linear combination of non linear functions of V_j s. This facilitates the computation of the law of the effective field resulting from the limit $N \to \infty$.

6.3.3) Is there an homeomorphism mapping the microscopic system to the mean field one?

We now return to this question, raised in the introduction. In our opinion the context of mean-field theory in neuronal network offers indeed an interesting context to study possible instanciations of this general question. Clearly, even in this context it would deserve a thorough investigation. What can be said, here, as a conclusion of this lecture, is that naive mean-field equations give somewhat an explicit example where the structure of the mesoscopic equations is the same as the microscopic ones (compare eq. (3) and (5).

Now, the more realistic case of random synaptic weights opens up the question of generalisation of this property. Clearly, even in the simplest case of independent couplings and one population, the mesoscopic equations have a different structure as the microscopic, due to the presence of time-dependent fluctuations (covariance) precisely induced by the variability on J_{ij}. The equation ruling the covariance, (13), introduce a clear complexification in the analysis and in the interpretation of solutions. More generally, it raises the critical question of relevant order parameters and their dynamics when averaging microscopic solutions.

$$\int \frac{dk}{2\pi}\frac{dk'}{2\pi}\hat{f}(k)\hat{f}(k')\exp -\left\{\frac{g^2}{2}\Delta(0)(k^2+k'^2) - g^2 k k'\Delta(t)\right\}$$

$$= \int_{-\infty}^{+\infty}\frac{dx}{\sqrt{2\pi}}e^{-x^2/2} f\Big(\underbrace{g\sqrt{\Delta(0)-|\Delta(t)|}\,x + g\sqrt{|\Delta(t)|}\,z}_{u}\Big) \; ; \qquad x = \frac{u - g\sqrt{\Delta(t)}\,z}{g\sqrt{\Delta(0)-\Delta(t)}}$$

$$= \int_{-\infty}^{+\infty}\frac{du}{g\sqrt{2\pi}\sqrt{\Delta(0)-\Delta(t)}}\, e^{-\frac{(u-g\sqrt{\Delta(t)}z)^2}{2g^2(\Delta(0)-\Delta(t))}} f(u)$$

$$\Gamma^2 = \int_{-\infty}^{+\infty}\frac{du\,dv}{2\pi g(\Delta(0)-\Delta(t))} f(u) f(v)\, e^{-\frac{(u-g\sqrt{\Delta(t)}z)^2}{2(\Delta(0)-\Delta(t))g^2}}\, e^{-\frac{(v-g\sqrt{\Delta(t)}z)^2}{2(\Delta(0)-\Delta(t))g^2}}$$

$$\Gamma^2 D_z = \int_{-\infty}^{+\infty}\frac{du\,dv}{2\pi(\Delta(0)-\Delta(t))g} f(u) f(v)\int\frac{dz}{\sqrt{2\pi}}e^{-z^2/2}\, e^{-\frac{(u-g\sqrt{\Delta(t)}z)^2}{2(\Delta(0)-\Delta(t))g^2}}\, e^{-\frac{(v-g\sqrt{\Delta(t)}z)^2}{2(\Delta(0)-\Delta(t))g^2}}$$

$$z^2 + \frac{(u-g\sqrt{\Delta(t)}z)^2}{g^2(\Delta(0)-\Delta(t))} + \frac{(v-g\sqrt{\Delta(t)}z)^2}{g^2(\Delta(0)-\Delta(t))}$$

$$\frac{g^2(\Delta(0)-\cancel{\Delta(t)})z^2 + u^2 - 2g\sqrt{\Delta(t)}\,uz + g^2\cancel{\Delta(t)}z^2 + v^2 - 2g\sqrt{\Delta(t)}\,z + g^2\cancel{\Delta(t)}z^2}{g^2(\Delta(0)-\Delta(t))}$$

$$= \frac{z^2\left[g^2(\Delta(0)-\Delta(t)) + 2g^2\Delta(t)\right] - 2g\sqrt{\Delta(t)}(u+v)z + u^2+v^2}{g^2(\Delta(0)-\Delta(t))}$$

$$= \left(z^2\left[g^2(\Delta(0)+\Delta(t))\right] - 2g\sqrt{\Delta(t)}(u+v)z + u^2+v^2\right)/g^2(\Delta(0)-\Delta(t))$$

$$= \left(g\sqrt{\Delta(0)+\Delta(t)}\,z - \frac{\sqrt{\Delta(t)}(u+v)}{\sqrt{\Delta(0)+\Delta(t)}}\right)^2 - \frac{\Delta(t)(u+v)^2}{(\Delta(0)+\Delta(t))g^2(\Delta(0)-\Delta(t))} + \frac{u^2+v^2}{g^2(\Delta(0)-\Delta(t))}$$

$$= \underbrace{\frac{\left(g\sqrt{\Delta(0)+\Delta(t)}\,z - \frac{\sqrt{\Delta(t)}(u+v)}{\sqrt{\Delta(0)+\Delta(t)}}\right)^2}{\left(g\sqrt{\Delta(0)-\Delta(t)}\right)^2}}_{X^2} \quad \underbrace{\frac{-\Delta(t)(u^2+v^2+2uv) + (\Delta(0)+\Delta(t))(u^2+v^2)}{g^2(\Delta(0)^2-\Delta^2(t))}} = Y$$

$$I^2 d_z = \int_{-\infty}^{+\infty} \frac{du\, dv}{2\pi g\,(\Delta(0)-\Delta(t))} \exp -\frac{1}{2}\, \frac{\Delta(0)(u^2+v^2)-2uv\,\Delta(t)}{g^2(\Delta^2(0)-\Delta^2(t))}$$

$$\times \int_{-\infty}^{+\infty} e^{-x^2/2}\, \frac{dz}{\sqrt{2\pi}} \quad ; \quad \cancel{dx = \frac{dz}{g\sqrt{\Delta(0)-2\Delta(t)}}}$$

$$dx = \frac{g\sqrt{\Delta(0)+\Delta(t)}\; dz}{g^2\sqrt{\Delta(0)-\Delta(t)}}$$

$$\Rightarrow$$

$${}^2 d_z = \int_{-\infty}^{+\infty} \frac{du\, dv}{2\pi}\, \frac{g^2\sqrt{\Delta(0)-\Delta(t)}}{g^2(\Delta(0)-\Delta(t))\sqrt{\Delta(0)+\Delta(t)}}\; e^{-\frac{1}{2}\frac{\Delta(0)(u^2+v^2)-2uv\Delta(t)}{g^2(\Delta^2(0)-\Delta^2(t))}}\, f(u)f(v)$$

$$\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}}\, \frac{dx}{\sqrt{2\pi}}$$

$$= \int_{-\infty}^{+\infty} \frac{du\, dv}{2\pi\sqrt{\Delta^2(0)-\Delta^2(t)}}\; e^{-\frac{1}{2}\frac{\Delta(0)(u^2+v^2)-2uv\Delta(t)}{g^2(\Delta^2(0)-\Delta^2(t))}}\, f(u)f(v)$$